A genomewide catalogue of single nucleotide polymorphisms in white-beaked and Atlantic white-sided dolphins

Fernandez, R; Schubert, M; Vargas-Velazquez, AM; Brownlow, A; Vikingsson, GA; Siebert, U; Jensen, LF; Oien, N; Wall, D; Rogan, E; Mikkelsen, B; Dabin, W; Alfarhan, AH; Alquraishi, SA; Al-Rasheid, KAS; Guillot, G; Orlando, L

Published in:
Molecular Ecology Resources

DOI:
10.1111/1755-0998.12427

First published: 07/05/2015

Document Version
Peer reviewed version

MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2016) **16**, 266–276 doi: 10.1111/1755-0998.12427

A genomewide catalogue of single nucleotide polymorphisms in white-beaked and Atlantic white-sided dolphins

R. FERNÁNDEZ,[1,†] M. SCHUBERT,[1] A. M. VARGAS-VELÁZQUEZ,[1] A. BROWNLOW,[2] G. A. VÍKINGSSON,[3] U. SIEBERT,[4] L. F. JENSEN,[5] N. ØIEN,[6] D. WALL,[7] E. ROGAN,[8] B. MIKKELSEN,[9] W. DABIN,[10] A. H. ALFARHAN,[11] S. A. ALQURAISHI,[11] K. A. S. AL-RASHEID,[11] G. GUILLOT[12] and L. ORLANDO[1]

[1]*Centre for GeoGenetics, University of Copenhagen, Øster Volgade 5-7, 1350K Copenhagen, Denmark,* [2]*Wildlife Unit, SAC Veterinary Services, Drummondhill, Stratherrick Road, Inverness, UK,* [3]*Marine Research Institute, Skulagata 4, 101 Reykjavík, Iceland,* [4]*Institute for Terrestrial and Aquatic Wildlife Research, University of Veterinary Medicine Hannover, Foundation, Werftstraße 6, Büsum, Germany,* [5]*Fisheries and Maritime Museum, Tarphagevej 2, Esbjerg, Denmark,* [6]*Institute for Marine Research, 5817, Bergen, Norway,* [7]*Irish Whale and Dolphin Group, Merchants Quay, Kilrush, Co., Clare, Ireland,* [8]*School of Biological, Earth and Environmental Sciences, University College Cork, Distillery Fields, North Mall, Cork, Ireland,* [9]*Natural History Museum, V. U. Hammersheimsgøta 13, 100 Tórshavn, Faroe Islands,* [10]*Centre de Recherche sur les mammiféres marins, Université La Rochelle, 5 allée de l'Océan, La Rochelle, France,* [11]*Zoology Department, College of Science, King Saud University, Riyadh 11451, Saudi Arabia,* [12]*Department of Applied Mathematics and Computer Science, Technical University of Denmark, Richard Petersens Plads, Lyngvy, Denmark*

Abstract

The field of population genetics is rapidly moving into population genomics as the quantity of data generated by high-throughput sequencing platforms increases. In this study, we used restriction-site-associated DNA sequencing (RADSeq) to recover genomewide genotypes from 70 white-beaked (*Lagenorhynchus albirostris*) and 43 Atlantic white-sided dolphins (*L. acutus*) gathered throughout their north-east Atlantic distribution range. Both species are at a high risk of being negatively affected by climate change. Here, we provide a resource of 38 240 RAD-tags and 52 981 nuclear SNPs shared between both species. We have estimated overall higher levels of nucleotide diversity in white-sided (π = 0.0492 ± 0.0006%) than in white-beaked dolphins (π = 0.0300 ± 0.0004%). White-sided dolphins sampled in the Faroe Islands, belonging to two pods (N = 7 and N = 11), showed similar levels of diversity (π = 0.0317 ± 0.0007% and 0.0267 ± 0.0006%, respectively) compared to unrelated individuals of the same species sampled elsewhere (e.g. π = 0.0285 ± 0.0007% for 11 Scottish individuals). No evidence of higher levels of kinship within pods can be derived from our analyses. When identifying the most likely number of genetic clusters among our sample set, we obtained an estimate of two to four clusters, corresponding to both species and possibly, two further clusters within each species. A higher diversity and lower population structuring was encountered in white-sided dolphins from the north-east Atlantic, in line with their preference for pelagic waters, as opposed to white-beaked dolphins that have a more patchy distribution, mainly across continental shelves.

Keywords: cetacean, *Lagenorhynchus*, population genomics, RADSeq, SNP

Received 8 February 2015; revision received 1 May 2015; accepted 5 May 2015

Introduction

Ecological and evolutionary studies of natural populations are increasingly based on molecular data at a genomewide scale (e.g. Emerson *et al.* 2010; McCormack *et al.* 2012; Reitzel *et al.* 2013; Fabre *et al.* 2014). Decreasing experimental costs related to the advent of high-throughput sequencing (HTS) methods make affordable the discovery, sequencing and genotyping of thousands of genetic markers, even for non-model organisms and taxa where little genetic information is available a priori (Davey *et al.* 2011).

In recent years, restriction-site-associated DNA sequencing approaches have become extremely popular

Correspondence: Ruth Fernández, Fax: +45-3393-4215;
E-mail: Ruth.Fernandez@ices.dk
†Present address: International Council for the Exploration of the Sea, HC Andersens Boulevard 44-46, Copenhagen, Denmark

(RADSeq; Davey & Blaxter 2011; Wang *et al.* 2012; Elshire *et al.* 2011; Peterson *et al.* 2012). This approach can deliver high-resolution population genomic data on thousands of genetic markers (single nucleotide polymorphism (SNPs)). With RADSeq, only short genomic regions located in the vicinity of enzymatic restriction sites are targeted and deep-sequenced, which considerably reduces sequencing costs. RADSeq is particularly appropriate when no closely related reference sequence is available and, therefore, when efficient capture probes targeting specific regions of interest are difficult to design based on the sequence information from only distant evolutionary relatives (Davey *et al.* 2011). The question of whether or not RADSeq, or other reduced-representation methods (Poland *et al.* 2012), will be replaced by whole-genome sequencing in conservation genetics in the near future is still contentious, especially given that many research questions can be addressed with only a limited set of genetic markers and that full-genome sequencing still remains prohibitive for most laboratories (Allendorf *et al.* 2010; Davey *et al.* 2011).

In this study, we used RADSeq for the discovery and genotyping of genetic markers in two related species of cetaceans: the white-beaked (*Lageno-rhynchus albirostris*) and Atlantic white-sided (*Lageno-rhynchus acutus*) dolphin. At present, only a limited number of complete genomes of cetacean species have been sequenced (Lindblad-Toh *et al.* 2011; Zhou *et al.* 2013; Yim *et al.* 2014; Foote *et al.* 2015), and only three (the bottlenose dolphin, *Tursiops truncatus*; the Yangtze River dolphin, *Lipotes vexillifer*; and the killer whale, *Orcinus orca*) belong to the same suborder as *Lagenorhynchus* (suborder Odontoceti; i.e. toothed whales). The bottlenose dolphin represents the closest relative to the two species targeted in this study that has been sequenced at the genome level. The most recent common ancestor between bottlenose and white-beaked dolphins is estimated to have lived 9.4 Ma, while that between bottlenose and white-sided dolphins probably lived around 8.9 Ma (McGowen *et al.* 2009). Given such divergent times and the fact that our samples had been stored over relatively long time periods (see Materials and methods section), RADSeq was preferred over target enrichment capture or RNA sequencing to characterize a genomewide catalogue of SNP markers that could help explore patterns of population differentiation over a significant part of the dolphins' geographical range.

White-beaked and white-sided dolphins are endemic to the North Atlantic (Shirihai & Jarrett 2006). Despite sharing the same climatic preferences, in the north-east (NE) Atlantic white-beaked and white-sided dolphins have different habitat preferences. While the former is restricted to shelf waters and divided by areas of deep oceanic waters into a number of geographically isolated aggregations, the latter currently spreads over a more continuous distribution in oceanic waters (Lambert 2012). Both species are at high risk of being negatively affected by climate change in the forthcoming decades, most likely through population fragmentation, with an up to 80% loss of suitable habitat predicted for the white-beaked dolphin, especially within its southernmost core range (MacLeod 2009; Lambert 2012; Lambert *et al.* 2014). Current climate change models predict a poleward habitat shift for both species (MacLeod 2009), following a pattern reminiscent of the ongoing shift observed for the closely related Pacific white-sided dolphin (*Lagenorhynchus obliquidens*; Salvadeo *et al.* 2010). Even though Atlantic white-sided and white-beaked dolphins are currently classified as 'Least Concern' by the International Union for the Conservation of Nature (IUCN), the 'Agreement on the Conservation of Small Cetaceans of the Baltic, North East Atlantic, Irish and North Seas' (ASCOBANS, 2009) considered the protection of white-beaked dolphins to be a priority for research on both sides of the North Atlantic together with the development of additional genetic studies, including new markers, for white-sided dolphins. The relevance of population studies addressing the demography and level of connectivity in both species is indisputable, especially in European waters that hold a significant part of their global populations (MacLeod 2013).

A previous phylogeographic survey identified three genetically differentiated populations (north-west (NW) Atlantic, north Norwegian and North Celtic Seas) among 116 white-beaked dolphins (Banguera-Hinestroza *et al.* 2010). The haplotype diversity was found to be moderate using 323 bp of the mitochondrial D-loop ($h = 0.7320 \pm 0.00031$). The nucleotide diversity was low ($\pi = 0.56 \pm 0.04\%$), possibly as a result of a recent demographic bottleneck resulting from human activities (ASCOBANS, 2009; Banguera-Hinestroza *et al.* 2010). In contrast, no clear substructure was found between NW ($N = 88$) and NE ($N = 267$) Atlantic white-sided dolphins (Banguera-Hinestroza *et al.* 2014). Within the NE Atlantic, genetic differentiation was significant between white-sided dolphins from the Faroe Islands and those from the North Sea and the Scottish coast when considering 322 bp of the mtDNA D-loop region. Microsatellite data, however, only revealed marginal differentiation between the populations in the Shetland Islands and the west coast of Scotland (Banguera-Hinestroza *et al.* 2014). Although overall haplotype diversity was high ($h = 0.9270 \pm 0.007$), nucleotide diversity was relatively low ($\pi = 0.9 \pm 0.3\%$), and there was evidence of historical bottlenecks and post-bottleneck expansions (Banguera-Hinestroza

et al. 2014). Mirimin *et al.* (2011) sequenced 599 bp of the D-loop region for 41 white-sided dolphins sampled in Ireland and reported similar levels of haplotype diversity ($h = 0.946 \pm 0.020$) and lower nucleotide diversity ($\pi = 0.7 \pm 0.4\%$) than the findings from Banguera-Hinestroza *et al.* (2014). Overall, population genetic information for both species is still preliminary and limited by the amount of genetic markers considered, which have been mostly mitochondrial. We therefore undertook a large-scale genetic survey of white-beaked and white-sided dolphins, focusing on the recovery of a large nuclear SNP panel of animals sampled across their NE Atlantic range. More specifically, we sampled 113 specimens of white-beaked ($N = 70$) and white-sided ($N = 43$) dolphins (Fig. 1) and used RADSeq to identify 52 981 SNPs in both species and 39 065 species-specific diagnostic markers. In addition, we explored the possible existence of intraspecific population substructure, approximated effective population sizes for both species and estimated genetic diversity indices. This data set will help define strategies for the future conservation of both species.

Materials and methods

RAD-tag sequencing

Samples (skin/muscle) were collected along the distribution range of the species within the NE Atlantic between 1978 and 2012. A total of 43 *L. acutus* samples (11 females, 32 males) were gathered from Ireland ($N = 12$), Scotland ($N = 11$), the Faroe Islands ($N = 18$), Germany (southern North Sea; $N = 1$) and France ($N = 1$), while 70 *L. albirostris* (40 females, 30 males) were sampled from Ireland ($N = 6$), Scotland ($N = 21$), Germany, Denmark and Norway (southern North Sea; $N = 18$), Norway (Barents Sea; $N = 6$), Iceland ($N = 18$) and France ($N = 1$; Table S1, Supporting information, Fig. 1). Samples were kept frozen at −20 °C, in 96% ethanol, or both. Samples were obtained from individuals that had become stranded on the coast, individuals by-caught in fishing operations or using remote biopsy sampling, except for specimens from the Faroe Islands that were hunted (NAMMCO Annual Report 2006) and belonged to three pods landed in three different days and in three different

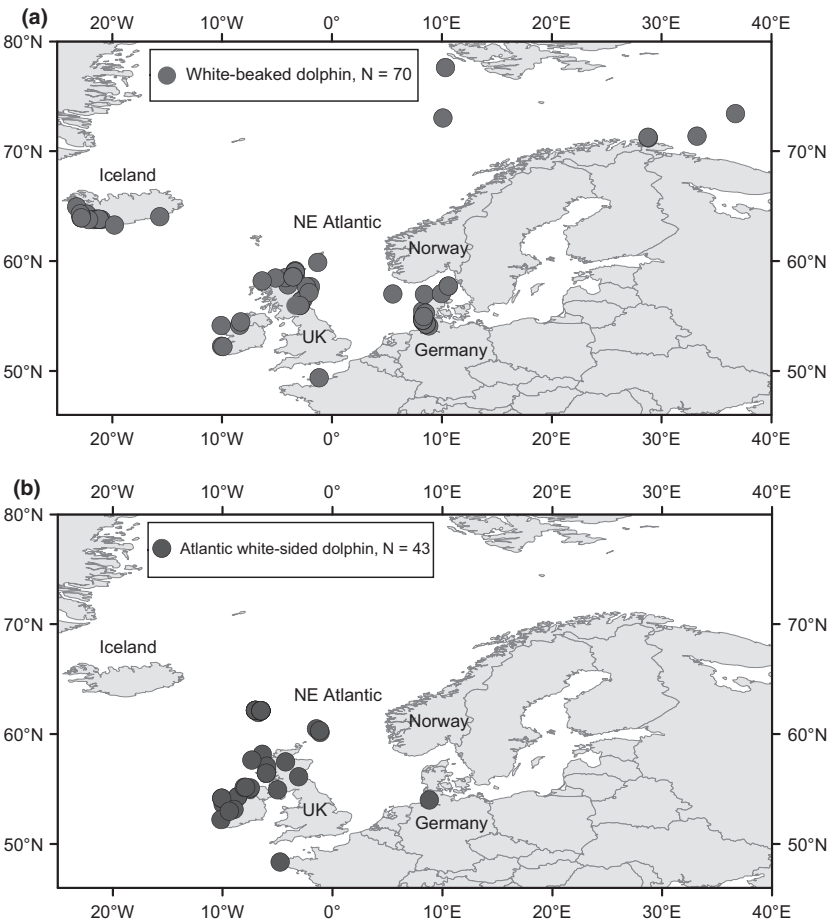

Fig. 1 Sampling locations of the specimens included in this study.

locations (Table S1, Supporting information). From one of the pods, only one individual was sequenced while seven and ten specimens from the other two pods were analysed. An additional *L. albirostris* sample from Ireland was processed and sequenced but not included in further analyses due to potential mislabelling of the sample.

Prior to DNA extraction, tissue was fragmented in a Tissuelyser (QIAGEN) using 0.5-mm DNA-free beads (QIAGEN). Genomic DNA was extracted using the DNeasy Blood and Tissue Kit (QIAGEN) following the manufacturer's instructions, except that samples were left digesting for a total of 14–38 h at 56 °C and that the elution buffer was incubated on the column at 37 °C for 15 min prior to final spinning. Blanks were included in all extractions to detect potential cross-contamination. Concentrations of DNA extracts used for library building were assessed using Qubit (Life Technologies) and are specified in Table S1 (Supporting information).

RADSeq libraries were built following a customized protocol based on Baird *et al.* (2008). In brief, 1 μg of genomic DNA was digested at 37 °C overnight for at least 14 h in a 50 μL reaction with 40–45 units of the restriction enzyme SbfI-HF (NEB). For each sample, the digested product was ligated to a P1 6-bp-barcoded adapter (Table S2, Supporting information) through a 60 μL reaction containing 1000 units of T4 ligase (NEB) and 1 mM rATP. Barcodes were selected to differ by a minimum of 2 nucleotides in order to minimize sample misclassification due to sequencing errors. After P1 adaptor ligation, libraries combining the products of four samples (2 μg total DNA) were built and sheared in a Bioruptor (Diagenode) using 4 cycles of 30 s on and 90 s off, with power settings on high. On one occasion, the library contained products of two samples only (GER_16411 and GER_3614). Whenever possible, samples with DNA extracts showing similar concentration were merged in the same library to avoid overrepresenting a particular sample. Fragmented products were size-selected to target 300- to 600-bp-long templates using Agencourt AMPure XP magnetic beads. The efficiency of the size-selection procedure was assessed using 2% agarose gels. Selected fragments were blunt ended (Quick blunting kit, NEB) and A-tailed (Klenow fragments, 3′ to 5′ exo, NEB) following manufacturer's instructions. A-tailed fragments were ligated to a P2 adapter (Table S2, Supporting information) through a 50 μL reaction containing 1000 units of T4 ligase (NEB) and 1 mM rATP.

PCRs of 25 μL total volume were carried out consisting of 1× AmpliTag Gold PCR buffer, 4 mM $MgCl_2$, 1 μg BSA, 2.5 units AmpliTaq Gold (Life Technologies), 1 mM dNTP, 0.5 μM of forward and reverse Illumina TruSeq primers (Table S2, Supporting information) and 14.25 μL library template. After an initial denaturation at 92 °C for 10 min, we performed 26–28 amplification cycles of 92 °C for 30 s, 60 °C for 30 s, 72 °C for 40 s and a final extension of 72 °C for 7 min. Libraries were pooled (8–10 libraries per lane) and sequenced as single reads on an Illumina HiSeq2000 platform at the Danish National High-throughput DNA Sequencing Centre.

QPCR sex determination

For the majority of the specimens ($N = 98$), sex was determined by qualified staff during necropsies. For 15 individuals, mostly those sampled through biopsies and for which information on gender was not available or unconfirmed, sex was determined by qPCR (Table S1, Supporting information) using the primers CetZFXY_F1 and CetZFXY_R1 and CETZFX and CETZFY probes based on Morin *et al.* (2005). PCRs of 30 μL total volume were carried out and consisted of 1 μL of DNA, 1× AmpliTaq Gold PCR buffer, 2.5 mM $MgCl_2$, 12 μg BSA, 1.2 units AmpliTaq Gold, 0.25 mM dNTP, 0.25 μM of forward and reverse primers and 0.2 μM of each probe. PCR amplification was performed in a Stratagene qPCR system under the following conditions: after an initial denaturation at 95 °C for 5 min, 60 amplification cycles of 95 °C for 10 s and 60 °C for 1 min were carried out followed by a final extension of 40 °C for 1 min.

SNP identification

Illumina reads were demultiplexed per library based on the sequence of primer indexes. The software ADAPTERREMOVAL (Lindgreen 2012) was used to trim potential adapter sequences at the 3′ end of the reads (Fig. S1, Supporting information). In addition, stretches of N's and bases with a Phred quality score lower than 2 were removed from the ends of the reads. Only reads equal to or longer than 36 bp following trimming were retained. Individual libraries were demultiplexed per sample (based on the sequence of the different P1 6-bp-barcoded adapters used), quality filtered and assembled using the STACKS tool kit v. 1.09 (Catchen *et al.* 2013). First, reads were demultiplexed based on their individual barcodes and quality filtered using process_radtags. No mismatches were tolerated within individual barcodes or RAD-tag cut sites, and we discarded sequences for which the mean raw Phred quality score dropped below 20 within a sliding window consisting of 15% of the read length. Preliminary analyses with no further trimming/filtering showed an increase in SNP frequency towards the 3′ end of the reads. As the latter probably corresponds to spurious polymorphisms resulting from increasing amounts of sequencing errors (Fig. S2, Supporting information), we retained reads truncated to 76 bp.

As no reference genome is available for *L. albirostris* and *L. acutus*, the ustacks, cstacks and sstacks options

within STACKS were used to de novo assemble loci. All the samples listed in Table S1 (Supporting information) were used to build a RAD-tag catalogue. The minimum number of identical reads (m) needed to form a stack was set to 3. For each individual, the maximum number of mismatches allowed to group stacks as a single locus was set to 2 (M). Loci identified in different individuals were merged into single loci, or RAD-tags, when a maximum of 2 mismatches was found (n). Prior to the selection of the most adequate final set of parameters, tests were run separately for five white-beaked and five white-sided dolphins, where we explored the use of Phred quality scores of 10, 'M' values between one and five, and 'n' values between zero and six. Overall, our selected parameter settings largely matched the recommendations from Viricel *et al.* (2014) when assessing the use of RAD-tag genotyping for interfamilial comparisons of two species of cetaceans. Highly repetitive sequences were removed or broken down using the '-t' option. Each RAD-tag was given an ID and included into a reference catalogue.

We subsequently applied the correction module rxstacks to the obtained data in order to (i) re-evaluate SNP calls in every individual using the bounded SNP model to recall the site as either homozygous or heterozygous, (ii) blacklist RAD-tags matched by several stacks for a single individual in the population, (iii) prune excessive haplotypes from individual loci according to the prevalence of haplotypes for a given catalogue locus and (iv) assign a likelihood to every locus in the population, which reflects the overall confidence of the SNP call. Finally, the *populations* module from STACKS was used to identify a final set of orthologous loci and SNPs. For any individual, loci were retained if the depth of coverage was at least five.

Data analysis

The *populations* module within STACKS was used to approximate heterozygosity and nucleotide diversity values. For such analyses, each species was considered as a differentiated population. RAD-tags showing more than two SNPs (N = 6187) were excluded from the analysis. We used the ADMIXTURE program (Alexander *et al.* 2009) to detect population structure, considering various numbers of putative clusters ranging from 1 to 6 and computing fivefold cross-validation error to select the most suitable number of clusters (Alexander & Lange 2011).

Results

Identification of RAD-tags

A total of 497 394 094 sequences were analysed through *process_radtags*. Of these, 152 438 128 were removed due to ambiguous barcodes (no mismatch allowed), 34 521 074 were removed due to ambiguous RAD-tag cut sites (no mismatch allowed), and 34 371 085 did not meet our quality (see Materials and methods section) or length (76 bp) criteria. Thus, a total of 276 063 807 reads of 76bp were retained as the input for subsequent components of the pipeline. Note that these numbers include the additional Irish sample that was processed and sequenced but not included in downstream analyses (see the Materials and methods section).

A total of 179 170 RAD-tags were identified among the 113 individuals analysed and stored as a RAD-tag catalogue. The average number of loci that matched the catalogue per individual was 15 395 (with individual counts ranging from 4 to 73 967). Average numbers were slightly higher for *L. albirostris* (15 703; with individual counts ranging from 4 to 73 967) than for *L. acutus* (14 893; with individual counts ranging from 637 to 73 124). From the RAD-tags present in the catalogue, 122 430 (68.3%) were monomorphic, 32 994 (18.4%) contained one SNP, and 17 559 (9.8%) contained two SNPs. From these, 38 240 RAD-tags with one or two SNPs shared between white-beaked and white-sided dolphins are provided here as a resource of 52 981 nuclear SNPs (Table S3, Supporting information). A total number of 39 065 SNPs were found to be species-specific diagnostic markers. The SNP density present in our resource RAD-tag catalogue is one SNP every 193 bp, and the transition/transversion ratio is 2.56 (Fig. 2). The number of SNPs per individual varied between zero (for the individual from which only 4 RAD-tags were recovered,

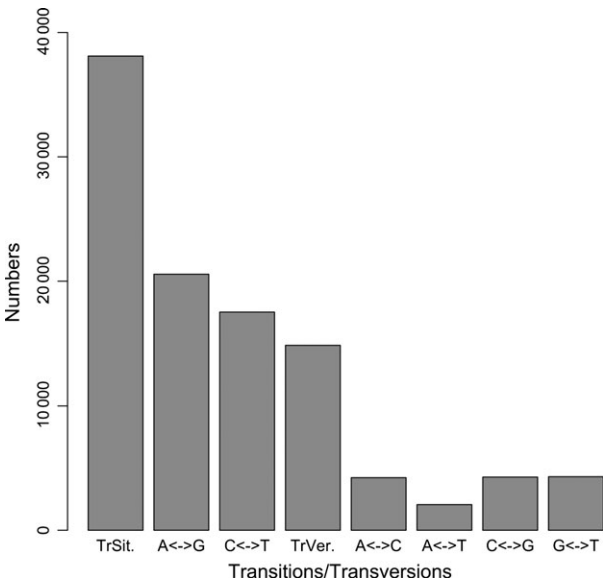

Fig. 2 Transitions and transversions occurring within the panel of 52 981 SNPs. TrSit = Transition; TrVer = Transversion.

IRE_117, Table S1, Supporting information) and 42 313 (for the individual FAE_394, Table S1, Supporting information).

Population structure

Genomewide levels of diversity were estimated through the *populations* module included in STACKS. *L. albirostris* showed consistently lower levels of genetic variation than *L. acutus*, with estimates corresponding to about 61% and 62% of the nucleotide diversity and heterozygosity observed in *L. acutus*, respectively (Table 1). Similarly to Pujolar *et al*. (2013), and assuming a neutral model of evolution and a random-mating population of constant size, effective population size (N_e) can be approximated using our estimates of nucleotide diversity (π) and the formula $\pi = 4*N_e*\mu$ (Tajima 1983), where μ is the mutation rate per site per generation. Considering estimates of mutation rates for the bottlenose dolphin, 0.84×10^{-9} mutations per site per generation, or for the Yangtze River dolphin, 1.22×10^{-9} (Zhou *et al*. 2013), our estimates of N_e ranged between 61 253 and 88 963 for *L. albirostris* and between 100 775 and 146 363 for *L. acutus*.

To fulfil the independence assumption required in ADMIXTURE analyses (Alexander *et al*. 2009), we randomly selected one SNP when a given RAD-tag showed two SNPs. Based on the cross-validation error, ADMIXTURE determined the most likely number of clusters among our sample set to be four ($K = 4$; Fig. 3), although cross-validation error values were very similar for $K = 2$ (0.427), $K = 3$ (0.430) and $K = 4$ (0.422; Fig. 3). This partition of the samples therefore established the presence of one or two clusters within each of the two species (Fig. 4a,b). For $K = 3$, all *L. acutus* are assigned to a single cluster, while *L. albirostris* (except ICE_SV2, see below) are assigned to two additional clusters (figure not shown). Note that only 112 dolphins have been included in this analysis because no SNPs were recorded for individual IRE_117.

Interestingly, and regardless of the value considered for *K*, the sample ICE_SV2, identified in its necropsy as *L. albirostris*, is genetically assigned to *L. acutus* (Fig. 4a,b). This result could represent a species misclassification of the individual in the field or a misclassification due to the admittedly low number of SNPs involved in our analysis for this individual (53 SNPs distributed along 42 loci). Alternatively, it could also indicate the presence of hybrids between both species. To further explore these hypotheses, we determined the number of species-diagnostic SNPs by carrying out an additional run of the *populations* module within STACKS where the potentially admixed individual was excluded from the analysis. When excluding specimen ICE_SV2, 154 664 loci and 39 086 species-specific SNPs were identified. We found 23 of these diagnostic SNPs in ICE_SV2 from which two SNPs showed alleles typical of *L. albirostris* (one locus being sampled in only three other *L. albirostris* and one *L. acutus*) and the remaining 21 loci showed alleles typical of *L. acutus* (Table S4, Supporting information).

To rule out the possibility of having misclassified ICE_SV2 because of its low coverage, we down-sampled the genotypes of all other individuals that shared at least 10 SNPs with ICE_SV2 ($N = 28$) so that only the 42 loci used for ICE_SV2 were considered in an additional ADMIXTURE run. ADMIXTURE results for $K = 2$ (two genetically distinct groups assumed) correctly assigned all individuals to their expected species except for ICE_SV2, which was assigned to the *L. acutus* group (Fig. S3, Supporting information).

Finally, we assessed the levels of heterozygosity and nucleotide diversity within two of the pods of *L. acutus* sampled in the Faroe Islands. Levels of diversity were lower when each of the pods was studied separately than when all Faroese dolphins were analysed together. However, indices were similar to those encountered in

Table 1 Diversity indices estimates for *L. albirostris* and *L. acutus* (% ± standard error). All positions (variant and fixed) are considered

Species	Nucleotide diversity, π	Observed heterozygosity
L. albirostris	0.0300 ± 0.0004	0.0120 ± 0.0002
L. acutus	0.0492 ± 0.0006	0.0195 ± 0.0004

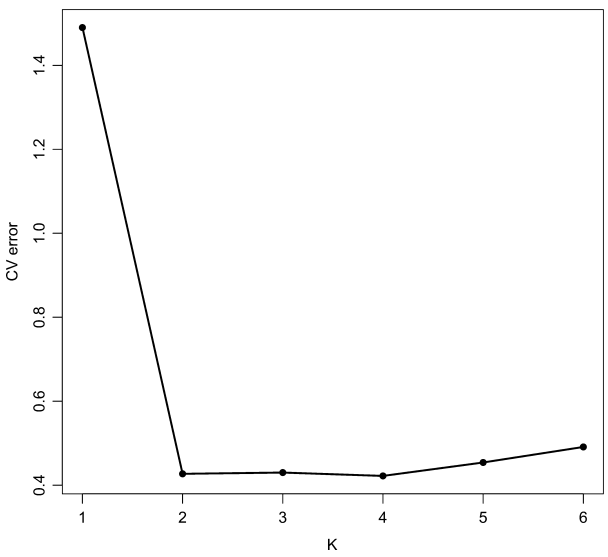

Fig. 3 Cross-validation (CV) error (*y* axis) for *K* between 1 and 6 as given by ADMIXTURE.

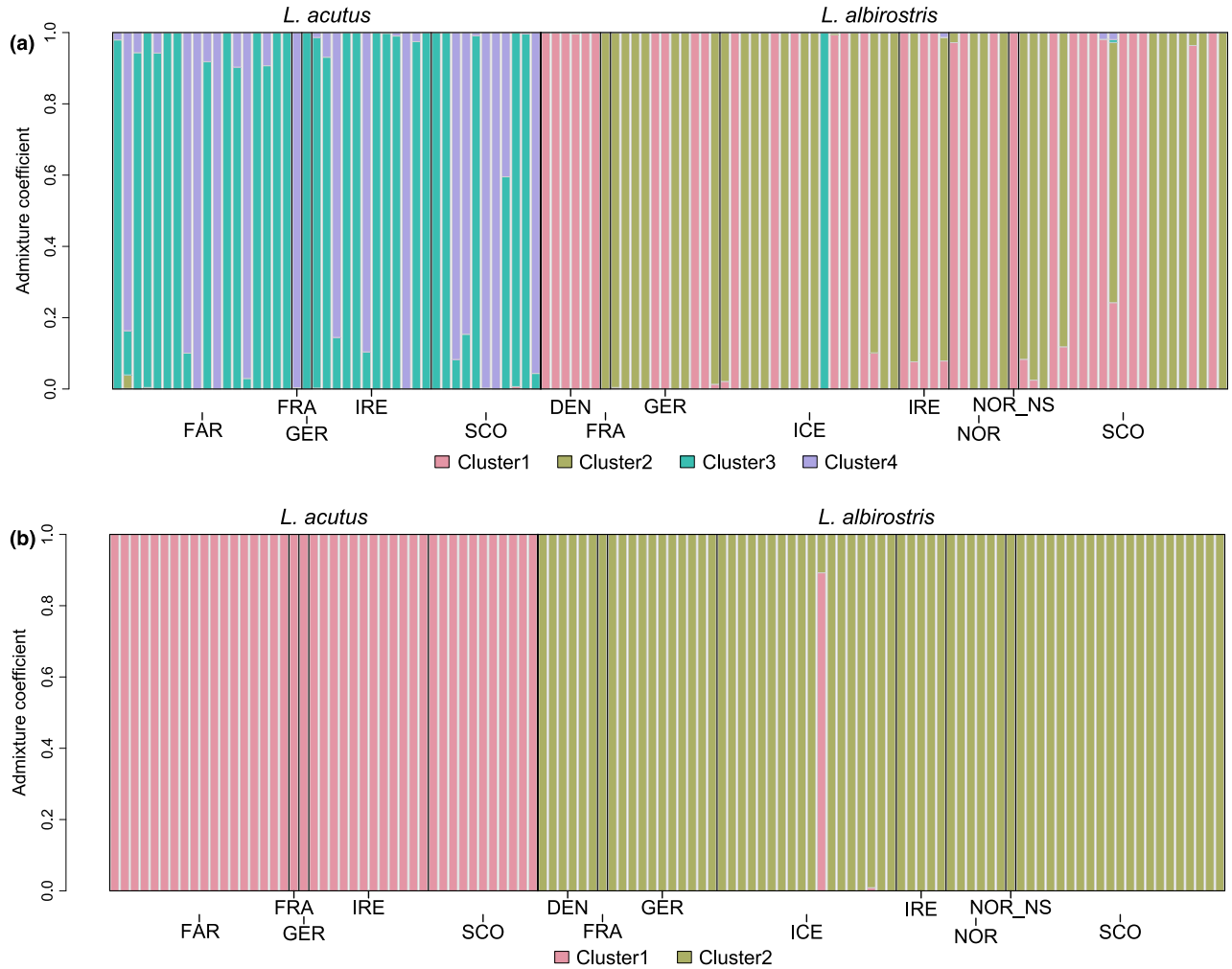

Fig. 4 ADMIXTURE results for 112 individuals, $K = 4$ (a) and $K = 2$ (b). FAR = Faroe Islands; FRA = France; GER = Germany; IRE = Ireland; SCO = Scotland; DEN = Denmark; ICE = Iceland; NOR = Norway; NOR-NS = Norway North Sea.

Table 2 Diversity indices estimates for *L. acutus* subgroups (% ± standard error). All positions (variant and fixed) are considered. * one specimen (FAE_394; see Table S1, Supporting information) collected on 05 September 2003 included

Subgroup	Nucleotide diversity, π	Observed heterozygosity
Faroe, pod collected on 26/08/2003, $N = 7$	0.0317 ± 0.0007	0.0225 ± 0.0006
Faroe, pod collected on 08/09/2004, $N = 10$	0.0267 ± 0.0006	0.0131 ± 0.0005
Faroe Islands all, $N = 18$*	0.0427 ± 0.0006	0.0219 ± 0.0005
Ireland, $N = 12$	0.0331 ± 0.0006	0.0127 ± 0.0004
Scotland, $N = 11$	0.0285 ± 0.0007	0.0154 ± 0.0006

larger geographical areas (i.e. Ireland, Scotland; Table 2) where a similar number of dolphins, sampled throughout an 11- to 14-year period, were studied.

Discussion

New panel of genetic markers

We present a large set of loci and SNPs discovered by RADSeq and shared between two related cetacean species. Previous population genetic studies of both species have sequenced fragments of the mtDNA control region (Banguera-Hinestroza *et al.* 2010, 2014; Mirimin *et al.* 2011) and <15 microsatellite markers, which were either specifically designed for the white-beaked dolphin (Banguera-Hinestroza *et al.* 2010, 2014) or were based on resources developed for other cetacean species (Mirimin *et al.* 2011; Banguera-Hinestroza *et al.* 2014). Only the white-beaked dolphin mitogenome has been published (GenBank Acc. No. AJ554061; Arnason *et al.*, 2004) and has been used in phylogenetic studies (i.e. Arnason *et al.*, 2004; Vilstrup *et al.* 2011). Hence, limited genetic

resources were available for the two species, particularly for the white-sided dolphin, and the novel set of 52 981 SNPs distributed along 38 240 shared RAD-tags between both dolphin species described here represents a major advance. Our SNP panel could be used in future studies of *Lagenorhynchus* spp., for example carrying out RAD-Seq of additional samples, selecting a set of markers to be used in SNP chip genotyping experiments and/or target enrichment across a larger number of individuals. The latter approach will offer the additional advantage of being compatible with the analysis of historical and ancient specimens preserved in museum collections (e.g. Tin *et al.*, 2014), which are often characterized by degraded DNA material, and will therefore allow investigations over a historical temporal range.

RADSeq methodology pros and cons

RADSeq constitutes a useful methodology for genomewide population genetic studies of non-model organisms by providing large panels of genomewide markers at relatively low cost. In our particular study, however, we found two major limitations with the methodology: failure to sequence the same loci for every sample of interest, which may result in large amounts of missing data for a given SNP, and our inability to remove PCR duplicates (as within RAD-tags, 5′-read coordinates all start at the same genomic location), which may result in an overrepresentation of a single allele and, therefore, an overestimation of homozygotes. The first of these limitations could have been caused by different starting DNA qualities and by DNA fragmentation that could have resulted from the failure of the enzyme to bind to every cut site. Although dependent on the choice of restriction enzymes, large genome sizes, such as those of dolphin species (i.e. for the bottlenose dolphin, a first genome draft of 2.49 Gb has been released; Lindblad-Toh *et al.* 2011), usually result in a high number of enzymatic cut sites (Davey *et al.* 2011) and with limited sequencing efforts, in allelic dropout. The second limitation can be overcome if paired-end sequencing is used, because PCR duplicates can then be identified using coordinates from both reads, thereby reducing the rate of false positives (Schubert *et al.* 2014). We used single-end sequencing in our study due to software limitations when working with non-model organisms that require same-length reads as input files in order to identify RAD-tags, alleles and, ultimately, SNPs (i.e. STACKS; Catchen *et al.* 2013). Additional software developments such as the proposed PYRAD pipeline (Eaton 2014) may also be able overcome this limitation.

We have used samples of animals kept in excellent preservation conditions, and we have monitored contamination and included blanks throughout the DNA extraction procedures and library building (see Materials and methods section). If a reference genome is available, however, the removal of misaligned reads would constitute a powerful additional measure to completely remove contamination due to exogenous DNA. Finally, several of our libraries were sequenced twice due to low coverage of one or several of the four dolphins included in the library after a first sequencing run. Resequencing provided additional reads that were used for RAD-tag and SNP calling. Sequencing new libraries of the same individuals instead of sequencing the same libraries twice could have provided a larger variability and, potentially, the sequencing of new RAD-tags flanking additional enzyme cut sites.

Population structure and biological findings

Our ADMIXTURE analyses highlighted a certain degree of population substructure in both species. Substructuring levels were higher for white-beaked dolphins than white-sided dolphins given that for an ADMIXTURE analysis assuming three genetic clusters, white-beaked showed substructuring, while white-sided were all assigned to the same deme. Population substructuring in both dolphins species could not be unequivocally linked to geographical region, given that individuals from the four identified clusters were found throughout the research areas (Fig. S4a–e, Supporting information). From this reason, and given the highly similar cross-validation errors obtained in ADMIXTURE, $K = 2$ may represent the most appropriate choice based on data from the current study.

Levels of nucleotide diversity were lower in white-beaked than in white-sided dolphins, in agreement with previous studies based on mtDNA control region sequences (Banguera-Hinestroza *et al.* 2010, 2014; Mirimin *et al.* 2011). We should caution that our genomewide diversity assessment is an underestimation because only RAD-tags containing 0–2 SNPs have been considered in the calculation. Our SNP density estimates, however, (one SNP every 193 bp when both species are considered together), are in line with findings from Viricel *et al.* (2014), who reported one SNP every 292 bp for five individuals of short-beaked common dolphins, *Delphinus delphis*. Our estimates are much higher than those suggested for the genetically depleted Bornean elephant, *Elephas maximus borneensis*, where sequence variability corresponded to approximately one SNP per 1250 bp for eight studied specimens (Sharma *et al.* 2012), which could reflect differences in mutation rates, generation times and other life history traits, as well as differences in their recent population history.

The Faroese government reported the capture of 312 and 617 white-sided dolphins in 2005 and 2006, respec-

tively (NAMMCO Annual Report 2006). The data on life history and abundance currently available for the two *Lagenorhynchus* species, however, are not believed to be sufficient to produce stock assessments and delineations (NAMMCO Annual Report 2010). Two white-sided dolphin pods sampled in the Faroe Islands showed levels of diversity similar to dolphin aggregates of similar sizes from other geographical areas that were sampled over a 11- to 14-year period (i.e. Scotland, Ireland). Thus, no evidence of higher levels of kinship within pods can be derived from our analyses. Mirimin *et al.* (2011) found that the white-sided dolphins sampled in two mass-stranding events ($N = 19$ and $N = 5$) in Ireland were mainly unrelated to each other and suggested that the species does not show natal philopatry. It should be noted that, in our case, only adult dolphins were sampled within each pod, which could have influenced the results and eliminated the possibility of having included mother–calf pairs in our analyses.

Based on our overall estimates of nucleotide diversity, estimates for long-term effective population sizes were calculated in each species (61 253–88 963 individuals for white-beaked and 100 775–146 363 for white-sided dolphins). These results should be taken with caution because the model used (see Materials and methods section) assumes both that sites evolve neutrally and an idealized random-mating population at a constant size, requirements that may not be fully met. Our findings, however, are in line with previous studies that have estimated the global population size of white-beaked dolphins (in the western and eastern parts of the North Atlantic) to lie between the high tens of thousands to low hundreds of thousands of individuals (Reeves *et al.* 1999). In addition, a survey covering the North Sea and adjacent waters provided an estimate of 7856 (coefficient variation, CV = 0.30) *L. albirostris* and 11 670 (CV = 0.26) *Lagenorhynchus* spp. in the research area (Hammond *et al.* 1995). A similar survey covering the same area in 2005 estimated an abundance of 16 536 (CV = 0.30) white-beaked dolphins (Hammond *et al.* 2013).

Specimen ICE_SV2 was by-caught in the gillnet fishery operating in Icelandic waters, and no pictures are available from this specimen. No phenotypic and/or colour pattern comparison of the dolphin is thus possible; hence, a potential misclassification of the species in the field cannot be ruled out. As opposed to white-beaked dolphins, white-sided dolphins are not common within Icelandic coastal waters (Víkingsson 2004; Pike *et al.* 2009), but sightings and strandings of the species have been reported within the area (e.g. JNCC, 2003, Víkingsson 2004). According to dedicated surveys, they are more abundant in offshore Icelandic waters (Sigurjónsson & Víkingsson 1997). Schaurich *et al.* (2012) reviewed 37 scientific publications between 1940 and 2010 and found a total of 27 hybridization events among cetaceans in the wild with several such episodes being recorded among dolphins ($N = 6$), including specimens of the genus *Lagenorhynchus* and close relatives *(i.e. L. obscurus* and *Delphinus capensis*, $N = 1$, and *L. obscurus* and *Lissodelphis peronii*, $N = 1$). It has been argued that cetacean hybridizations can occur towards the limits of the species' distribution ranges where reproductive resources are scarce and population sizes are small (Bérubé 2008). However, this does not seem to be the case for Icelandic white-beaked and white-sided dolphin populations because Icelandic coastal and offshore waters are well within the distribution ranges of both species (Shirihai & Jarrett 2006). Thus, although hybridization has been observed in a number of cetacean species both in captivity and in the wild, our findings require further investigation before it will be possible to evaluate the existence and importance of gene flow across these two species.

Conservation

Both of the species studied here are endemic to the North Atlantic and potentially threatened by ongoing and future climatic warming (Mac-Leod 2009; Lambert *et al.* 2014) as several parts of their distribution range are likely to become unsuitable within the next 100 years (Lambert *et al.* 2014). Protecting populations of white-beaked dolphins on both sides of the North Atlantic has been identified as a priority by ASCOBANS (2009). At the same time, ASCOBANS (2009) stresses the need to increase the availability of samples for genetic studies of both species. The comprehensive sample set and, most importantly, the panel of genomewide markers reported here represent the first step towards this goal and future conservation genetic studies.

Acknowledgements

We would like to thank Pernille Selmer Olsen and Tina Brand for help and advice in the laboratory; Andaine Seguin-Orlando, Kim Magnussen, Cecilie Mortensen and Lillian Petersen at the Danish National High-throughput DNA Sequencing Centre for their support with laboratory analyses; researchers from the Centre for GeoGenetics for discussions, especially the Paleomix group, Dr Anna-Sapfo Malaspinas and Dr Rute da Fonseca. We would like to thank Dr Julien Catchen, Dr Mark Rabinet and Dr Marti Pujolar for helpful advice in the analyses of RADSeq data. Thanks are due to Dr C.D. MacLeod, who encouraged R.F. to carry out this project and to Gema Hernández-Milián for helping with Irish samples. Irish samples were partially obtained from the Irish Cetacean Genetic Tissue Bank housed at the National Museum of Ireland – Natural History. Samples in Germany were collected through the stranding network funded by the State of Schleswig-Holstein, Germany. Dr R. Fernández was supported by a fellowship from the AXA Research fund. Mikkel Schubert was supported by Lundbeck Foundation grant

R52-A5062. This work was also supported by the Danish National Research Foundation (DNRF94), the Danish e-Infrastructure for Cooperation (DeIC), the International Research Group Program (IRG14-08), Deanship of Scientific Research (King Saud University, Saudi Arabia) and the U.S. National Institute for Mathematical and Biological Synthesis working group on Computational Landscape Genomics.

References

Alexander DH, Lange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, **12**, 246.

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics*, **11**, 697–709.

Arnason U, Gullberg A, Janke A (2004) Mitogenomic analyses provide new insights into cetacean origin and evolution. *Gene*, **333**, 27–34.

ASCOBANS (2009) Ascobans/Helcom small cetacean population structure workshop. 8-10 October 2007, Bonn, Germany.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Banguera-Hinestroza E, Bjørge A, Reid RJ, Jepson P, Hoelzel AR (2010) The influence of glacial epochs and habitat dependence on the diversity and phylogeography of a coastal dolphin species: *Lagenorhynchus albirostris*. *Conservation Genetics*, **11**, 1823–1836.

Banguera-Hinestroza E, Evans PGH, Mirimin L *et al.* (2014) Phylogeography and population dynamics of the white-sided dolphin (*Lagenorhynchus acutus*) in the North Atlantic. *Conservation Genetics*, **15**, 789–802.

Bérubé M (2008) Hybridism. In: *Encyclopedia of Marine Mammals* (eds Perrin WF, Würsig B, Thewissen JGM), pp. 588–592. Academic Press, San Diego.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Davey JW, Blaxter ML (2011) RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, **9**, 416–423.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499.

Eaton DA (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.

Elshire RJ, Claubitz JC, Sun Q (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**, e19379.

Emerson KJ, Merz CR, Catchen JM *et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *PNAS*, **107**, 16196–16200.

Fabre PH, Vilstrup JT, Raghavan M *et al.* (2014) Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. *Biology Letters*, **10**, 20140266. doi:10.1098/rsbl.2014.0266.

Foote AD, Liu Y, Thomas GWC *et al.* (2015) Convergent evolution of the genomes of marine mammals. *Nature Genetics*, **47**, 272–275.

Hammond PS, Berggren P, Benke H *et al.* (1995) Abundance of harbor porpoise and other cetaceans in the North Sea and adjacent waters. *Journal of Applied Ecology*, **39**, 361–376.

Hammond PS, MacLeod K, Berggren P *et al.* (2013) Cetacean abundance and distribution in European Atlantic shelf waters to inform conservation and management. *Biological Conservation*, **164**, 107–122.

JNCC (2003) Atlas of cetacean distribution in north-west European waters (eds. J.B. Reid, P.G.H. Evans, S.P. Northridge), p. 77, Joint Nature Conservation Committee, UK.

Lambert E (2012) *Environmental change, cetacean distribution and wildlife tourism in Western Scotland*. PhD thesis, University of Aberdeen, UK.

Lambert E, Pierce GJ, Hall K *et al.* (2014) Cetacean range and climate in the eastern North Atlantic: future predictions and implications for conservation. *Global Change Biology*, **20**, 1782–1793.

Lindblad-Toh K, Garber M, Zuk O *et al.* (2011) A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, **478**, 476–482.

Lindgreen S (2012) AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Research Notes*, **5**, 337.

MacLeod CD (2009) Global climate change, range changes and potential implications for the conservation of marine cetaceans: a review and synthesis. *Endangered Species Research*, **7**, 125–136.

MacLeod CD (2013) White-beaked dolphins in the Northeast Atlantic: a brief review of their ecology and potential threats to conservation status. In: *Towards a Conservation Strategy for white-beaked dolphins in the Northeast Atlantic* (eds MJ Tetley, SJ Dolman), pp. 26–35, Report from the European Cetacean Society's 27th Annual Conference Workshop, The Casa da Baía, Setúbal, Portugal. European Cetacean Society Special Publication Series No. XX, 121pp.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2012) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.

McGowen MR, Spaulding M, Gatesy J (2009) Divergence date estimation and comprehensive molecular tree of extant cetaceans. *Molecular Phylogenetics and Evolution*, **53**, 891–906.

Mirimin L, Banguera-Hinestroza E, Dillane E, Hoelzel AR, Cross TF, Rogan E (2011) Insights into genetic diversity, parentage, and group composition of Atlantic white-sided dolphins (*Lagenorhynchus acutus*) off the West of Ireland based on nuclear and mitochondrial genetic markers. *Journal of the Heredity*, **102**, 79–87.

Morin PA, Nestler A, Rubio-Cisneros NT, Robertson KM, Mesnick SL (2005) Interfamilial characterization of a region of the ZFX and ZFY genes facilitates sex determination in cetaceans and other mammals. *Molecular Ecology*, **14**, 3275–3286.

NAMMCO Annual Report 2006-Volume II. North Atlantic Marine Mammal Commission. 287 pp., Tromsø, Norway.

NAMMCO Annual Report 2010. North Atlantic Marine Mammal Commission. 501 pp., Tromsø, Norway.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double-digest RADSeq: an inexpensive method for *De Novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE*, **7**, e37135.

Pike DG, Paxton CGM, Gunnlaugsson T, Víkingsson GA (2009) Trends in the distribution and abundance of cetaceans from aerial surveys in Icelandic coastal waters, 1986-2001. *NAMMCO Scientific Publications*, **7**, 117–142.

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, **7**, e32253.

Pujolar JM, Jacobsen MW, Frydenberg J *et al.* (2013) A resource of genome-wide single-nucleotide polymorphisms generated by RAD-tag sequencing in the critically endangered European eel. *Molecular Ecology Resources*, **13**, 706–714.

Reeves R, Smeenk C, Kinze CC, Brownell RL, Lien J (1999) White-beaked dolphin *Lagenorhynchus albirostris*, Gray 1846. In: *Handbook of Marine Mammals* (eds Rodgway SH, Harrison RJ), pp. 1–30. Academic Press, London.

Reitzel AM, Herrera S, Layden MJ, Martindale MQ, Shank TM (2013) Going where traditional markers have not gone before: utility of a promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology*, **22**, 2953–2970.

Salvadeo CJ, Lluch-Belda D, Gómez-Gallardo A, Urbán-Ramírez J, MacLeod CD (2010) Climate change and a poleward shift in the distribution of the Pacific white-sided dolphin in the northeastern Pacific. *Endangered Species Research*, **11**, 13–19.

Schaurich MDN, Lopes FRV, de Oliveira LR (2012) Hybridization phenomenon in cetacean and pinniped species. *Neotropical Biology and Conservation*, **7**, 199–209.

Schubert M, Ermini L, Der Sarkissian C *et al.* (2014) Characterization of ancient and modern genomes by SNP detection and phylogenomic

and metagenomic analysis using PALEOMIX. *Nature Protocols*, **9**, 1056–1082.
Sharma R, Goossens B, Kun-Rodrigues C *et al.* (2012) Two different high throughput sequencing approaches identify thousands of de novo genomic markers for the genetically depleted Bornean elephant. *PLoS ONE*, **7**, e49533.
Shirihai H, Jarrett B (2006) *A Field Guide to the Marine Mammals of the World*. 384 pp, A&C Black, London.
Sigurjónsson J, Víkingsson GA (1997) Seasonal abundance of and estimated food consumption by cetaceans in Icelandic and adjacent waters. *Journal of Northwest Atlantic Fishery Science*, **22**, 271–287.
Tajima F (1983) Evolutionary relationships of DNA sequences in finite populations. *Genetics*, **105**, 437–460.
Tin MM-Y, Economo EP, Mikheyev AS (2014) Sequencing Degraded DNA from Non-Destructively Sampled Museum Specimens for RAD-Tagging and Low-Coverage Shotgun Phylogenetics. *PLoS ONE*, **9**, e96793.
Víkingsson GA (2004) Leiftur (White-sided dolphin). In: *Íslensk Spendýr (Mammals of Iceland)* (ed Hersteinsson P), pp. 158–159. Vaka-Helgafell, Reykjavik.
Vilstrup JT, Ho SYH, Foote AD *et al.* (2011) Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae. *BMC Evolutionary Biology*, **11**, 65.
Viricel A, Pante E, Dabin W, Simon-Bouhet B (2014) Applicability of RAD-tag genotyping for inter-familial comparisons: empirical data from two cetaceans. *Molecular Ecology Resources*, **14**, 597–605.
Wang S, Meyer E, McKay JK, Matz MV (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods*, **9**, 808–810.
Yim HS, Cho YS, Guang X *et al.* (2014) Minke whale genome and aquatic adaptations in cetaceans. *Nature Genetics*, **46**, 88–92.
Zhou X, Sun F, Xu S *et al.* (2013) Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptations. *Nature Communications*, **14**, 2708.

R.F. and L.O. planned and carried out the study. R.F., M.S., A.M.V., G.G. and L.O. conducted the bioinformatics and statistical analyses. A.B., G.A.V., U.S., L.F.J., N.Ø., D.W., E.R., B.M. and W.D. gathered and provided access to samples. A.H.A., S.A.A., K.A.S.A. and L.O. provided material and reagents. R.F. wrote the manuscript with contributions from all co-authors.

Data accessibility

Sequence reads have been deposited in the NCBI Sequence Read Archive (accession number SRP057668). The structure file used to run ADMIXTURE and a vcf file containing all SNPs recorded in the 113 sampled dolphins (see Table S1, Supporting information) have been stored in Dryad (doi:10.5061/dryad.sd28r).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Description of the specimens included in the present study (N = 113).

Table S2 Primers, P1 and P2 adapter sequences used in the present study.

Table S3 Text file encompassing all RAD-tags with one or two SNPs shared between white-beaked and white-sided dolphins.

Table S4 Species-specific diagnostic SNPs sampled in ICE_SV2.

Table S5 Polymorphic loci present in ICE_SV2.

Fig. S1 Length distribution of resulting reads after the use of ADAPTERREMOVAL (Lindgreen, 2012).

Fig. S2 Number of SNPs per nucleotide position (1–88).

Fig. S3 ADMIXTURE results for K = 2, 42 SNPs and 28 individuals.

Fig. S4 Clustering patterns of *L. acutus* (S4a) and *L. albirostris* (S4b) after the ADMIXTURE analysis for K = 4. A detailed view of clustering patterns of *L. albirostris* sampled across Iceland (S4c), Ireland and the UK (S4d) and Germany and Denmark (S4e) are provided.